

# Normalising multisensory cue reliabilities for individual participants in cue combination studies: A comparison of two methods

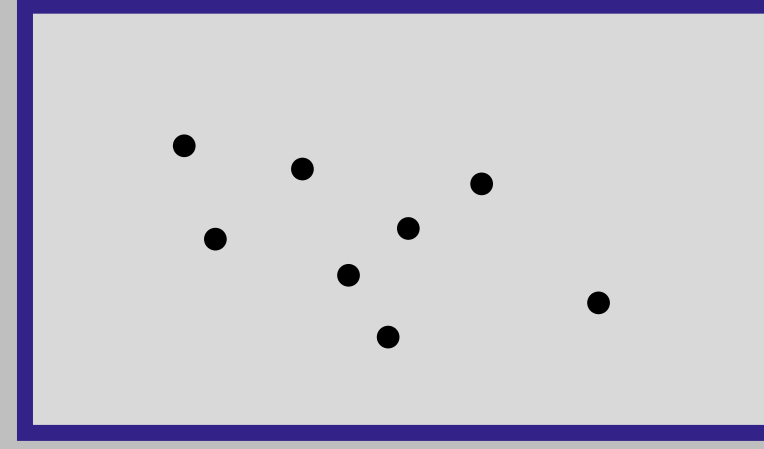
Sophia Hand, Thomas Chazelle, Sam Fenwick, Meike Scheller, Chris Allen, Marko Nardini

## Context of this study: A large experiment on individual differences in audiovisual integration

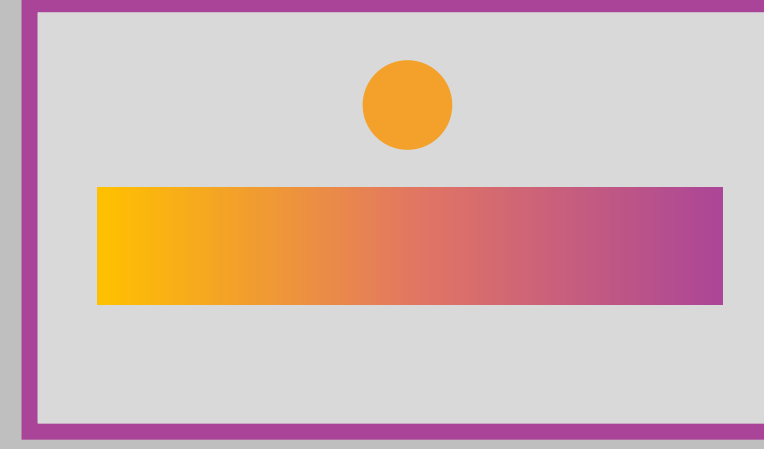
The ability to combine sensory cues together usually brings multisensory benefits: multisensory performance usually surpasses unisensory performance (Ernst & Banks, 2002). However, the tendency to integrate seems to vary across individuals and tasks, with previous studies finding little-to-no correlation between multisensory tasks within one same individual (Odegaard et al., 2016; Wilbiks et al., 2022).

In a large-N study (100 participants) and across two sessions, we are studying the stability and cue-specificity of multisensory benefits in a localisation task, using familiar and new cues.

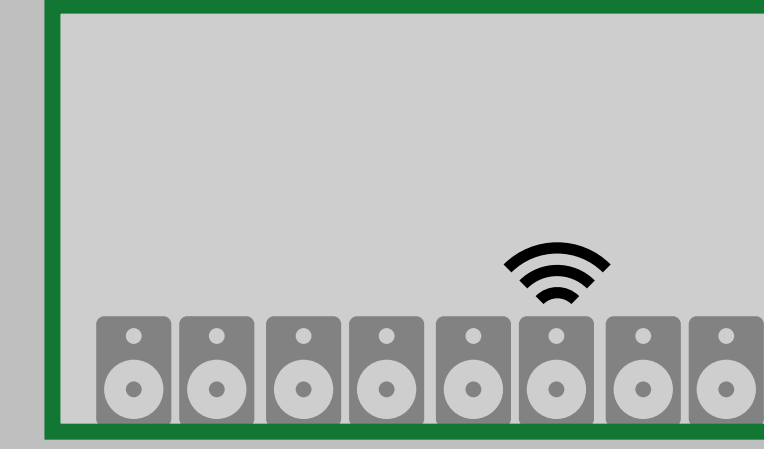
Spread - cloud of dots (familiar visual cue)



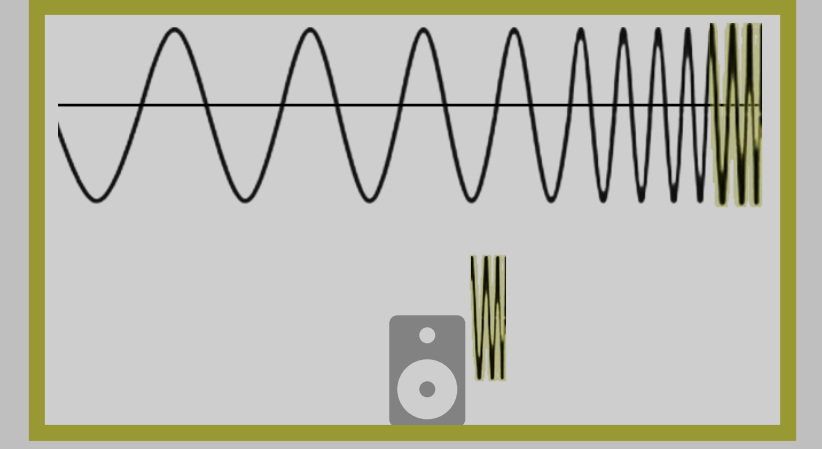
Colour (new visual cue)



Speaker position (familiar audio cue)



Pitch (new audio cue)

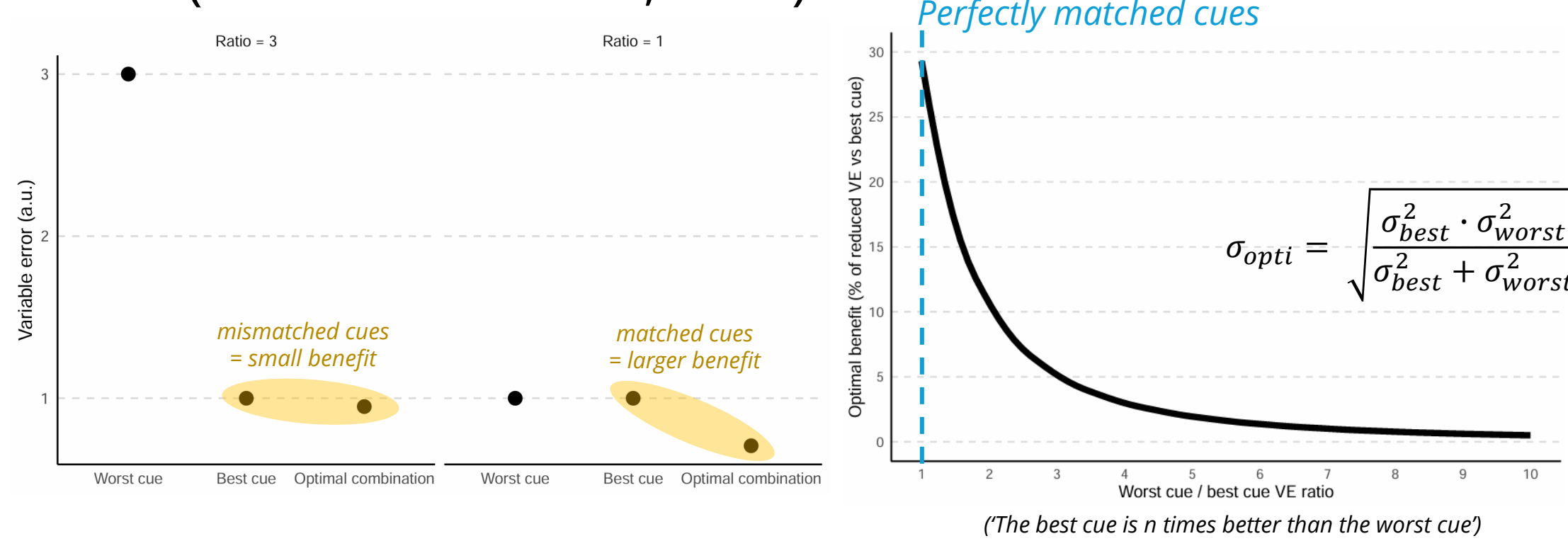


In the full experiment, participants performed a localisation task in which they adjusted a mouse cursor to show the location of a hidden target. The target could be shown by one or two out of four possible cues: spread, speaker location, as well as two new, arbitrary cues: colour and pitch.

## The importance of matching cues

One of the difficulties of measuring and comparing multisensory benefits is that **sensory cues are not equally helpful**. When integrating cues, it makes sense to give more weight to the best cues; this is called **optimal integration** (Ernst & Banks, 2002).

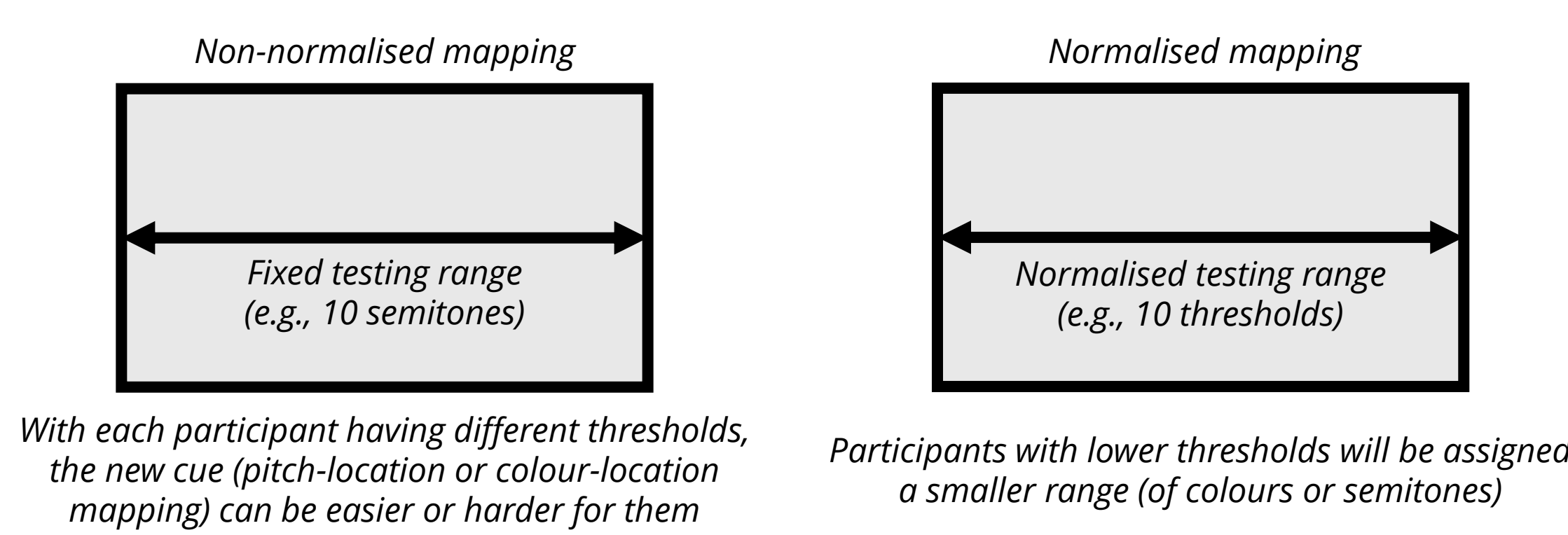
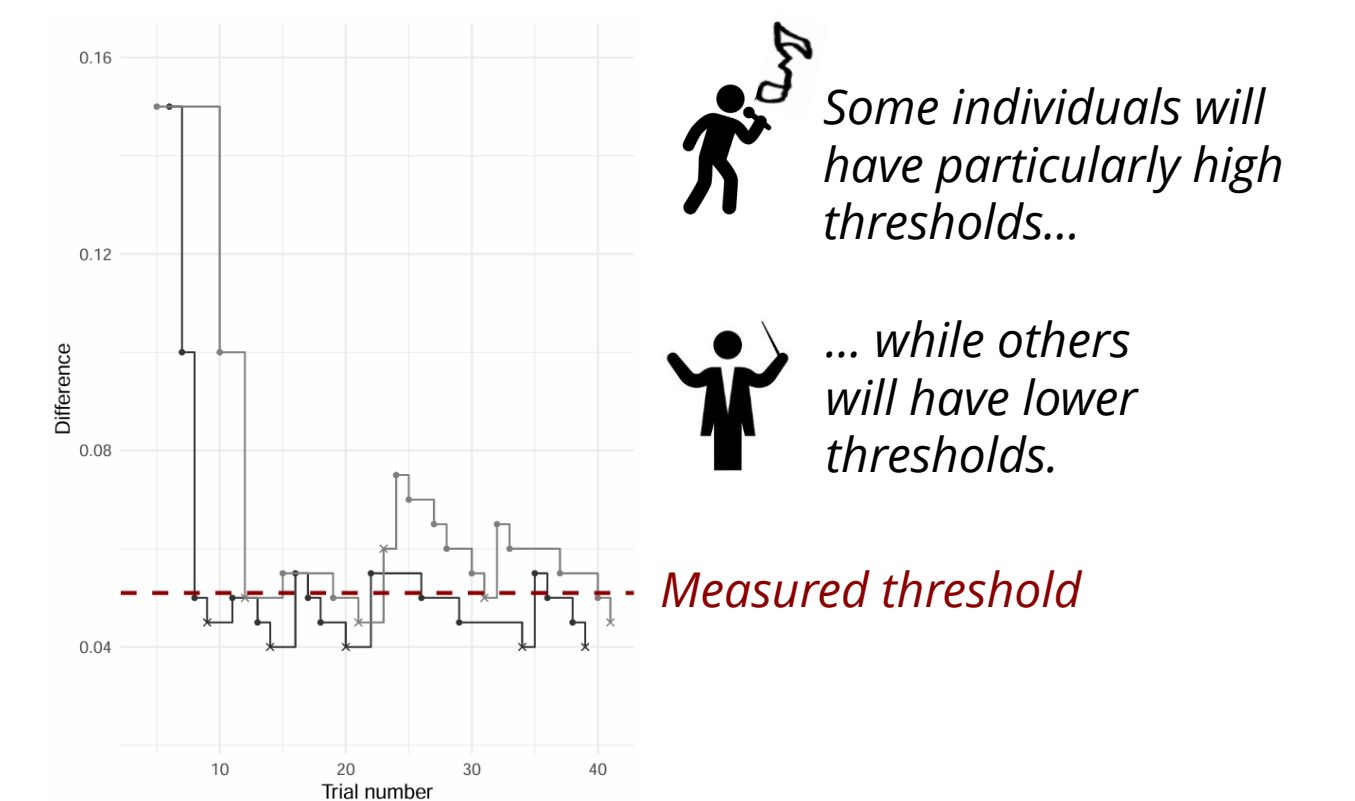
The optimal benefit from integration is larger when the cues have similar noise levels (variable error ratio = 1), but drops significantly for larger VE ratios, making the benefit harder to detect (Scheller & Nardini, 2024).



## Normalisation as a response to the problem of unmatched cues

The principle of normalisation is to tailor the difficulty of each cue to the participant's abilities.

A common approach to this consists in **first measuring a discrimination threshold** (e.g., a pitch discrimination threshold) to account for individual differences in the ability to use a cue (see on the right).

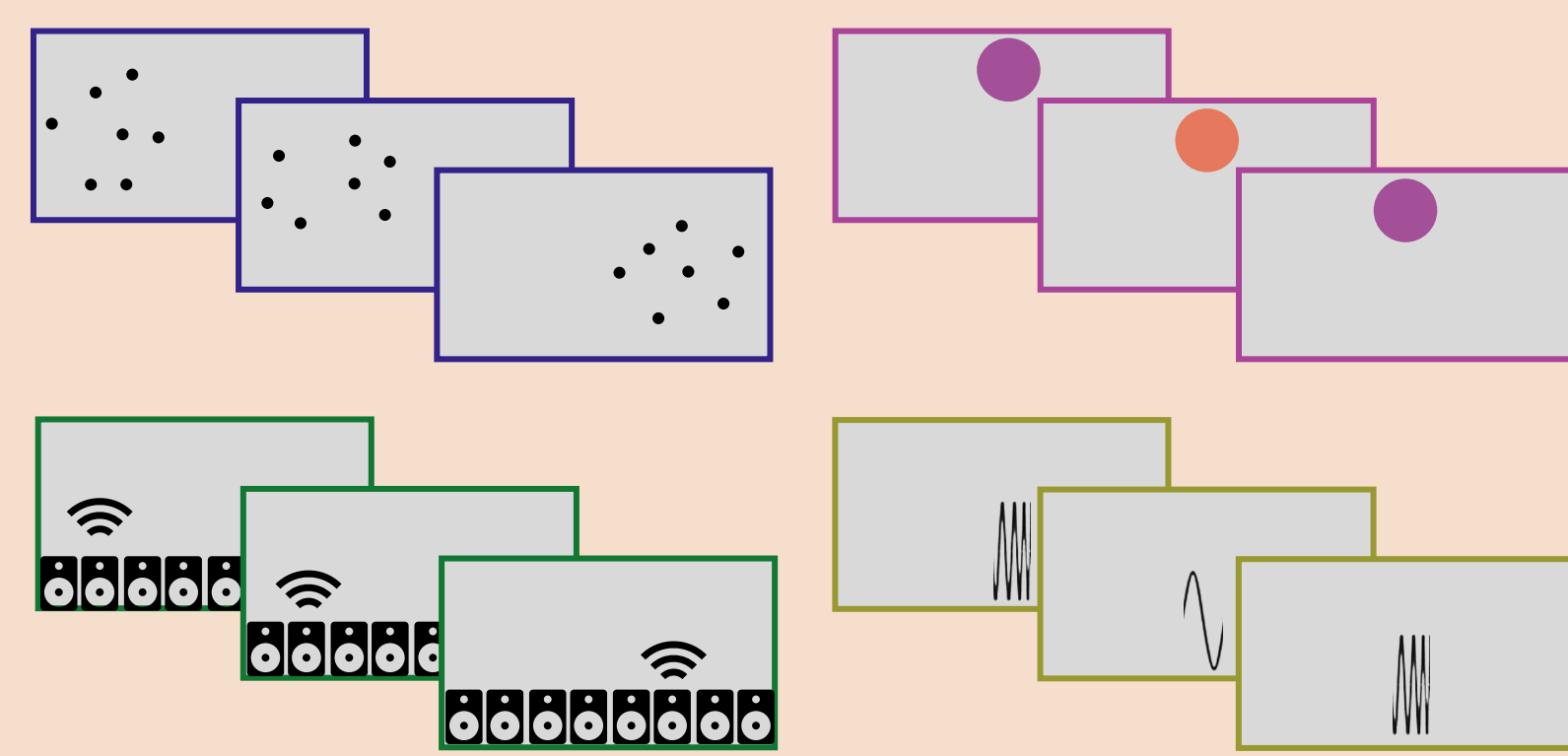


Once we obtain a discrimination threshold for each cue, we **use it to set the noise level in the cue** (e.g., define the pitch-space mapping; see on the left).

## Comparing two normalisation methods: does it matter how the threshold is determined?

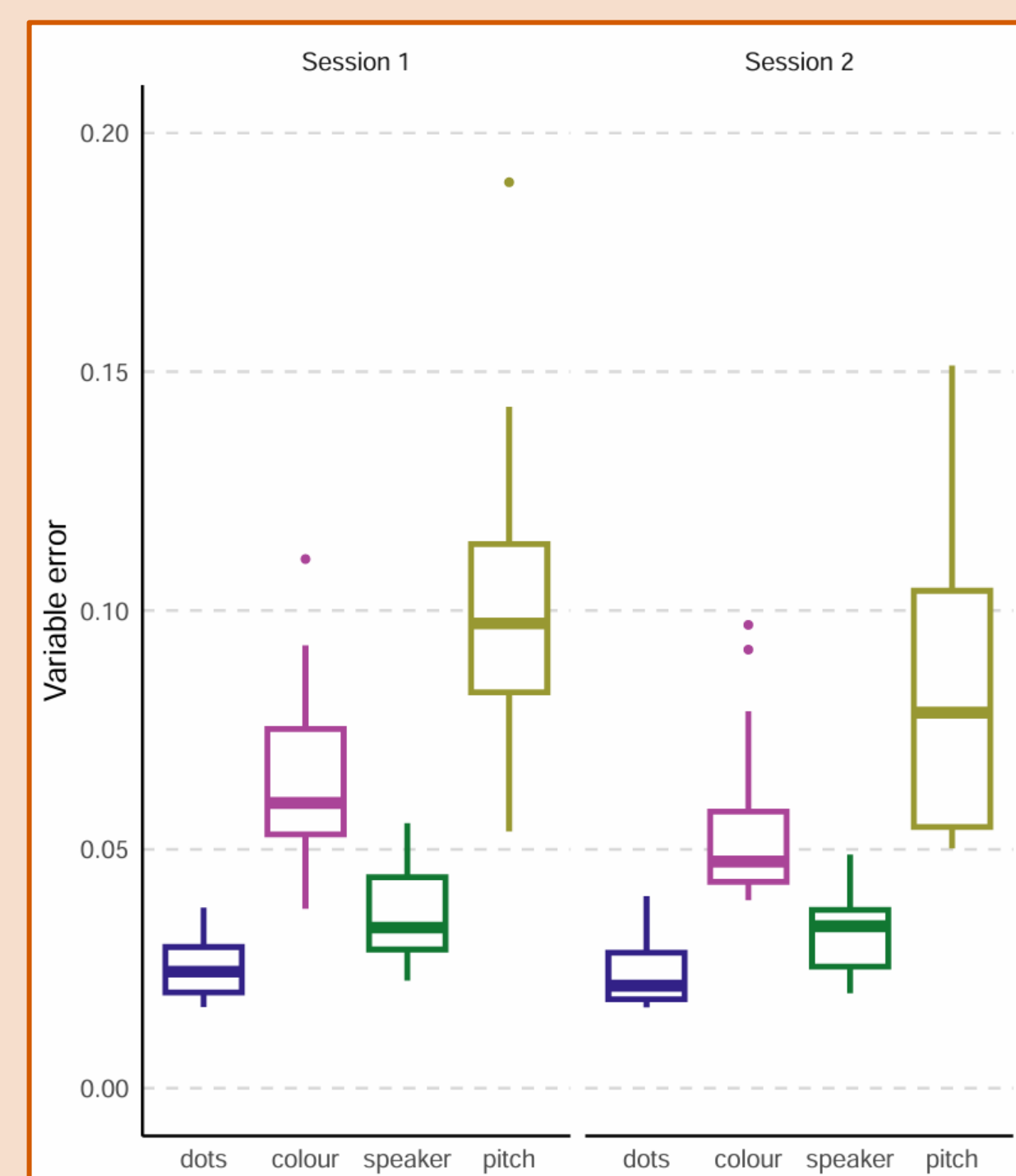
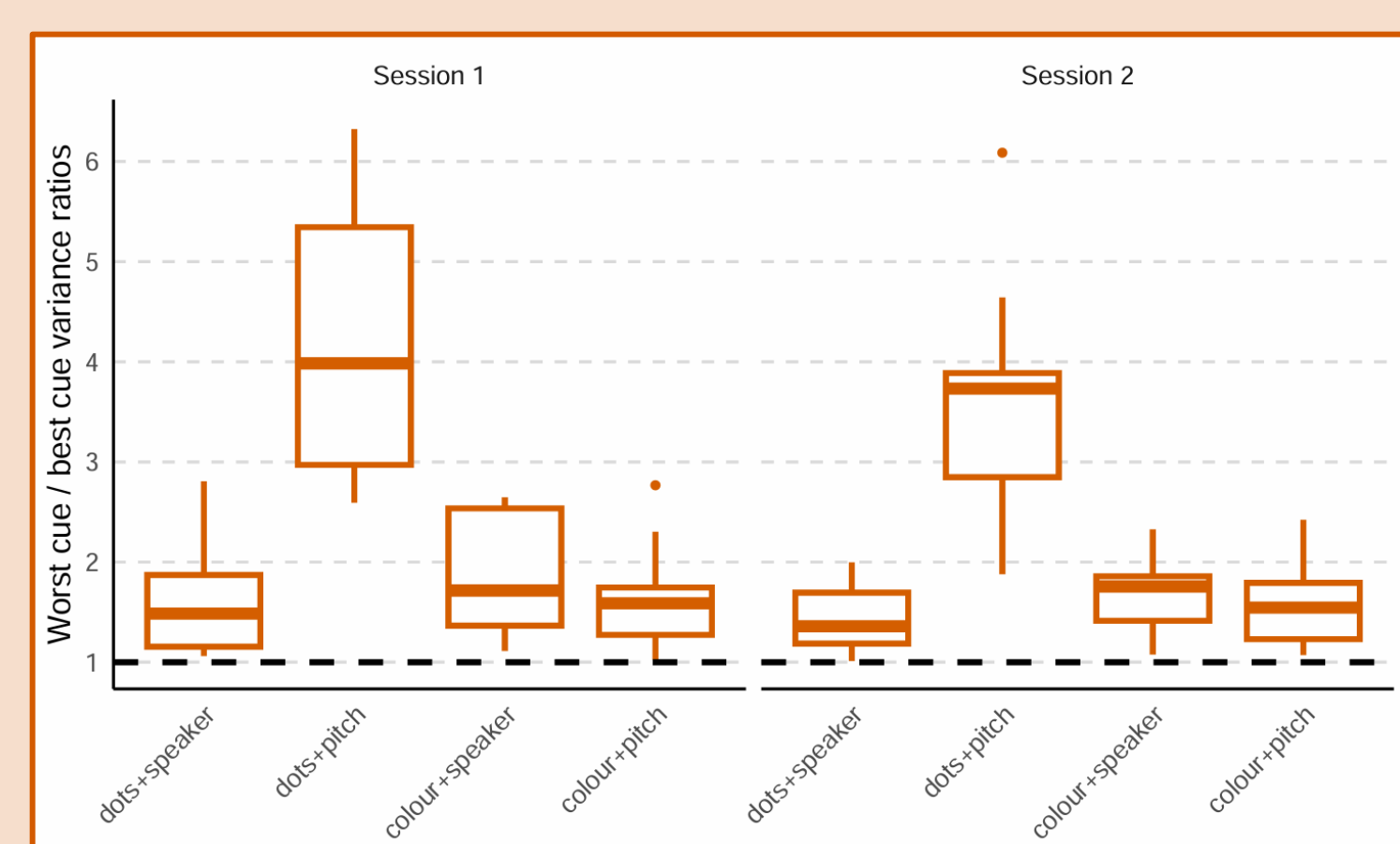
### Method 1: non-directional oddity task

In this task, 15 participants judged which of three stimuli differed from the other two. For each cue, we ran two staircases: the difference between the odd-one-out and the distractors got smaller (by 1 step) for each good answer and larger (by 2 steps) for each mistake (see above).



The thresholds obtained were then used to normalise colour and pitch in the localisation task: each cue should have the same number of thresholds in the testing range.

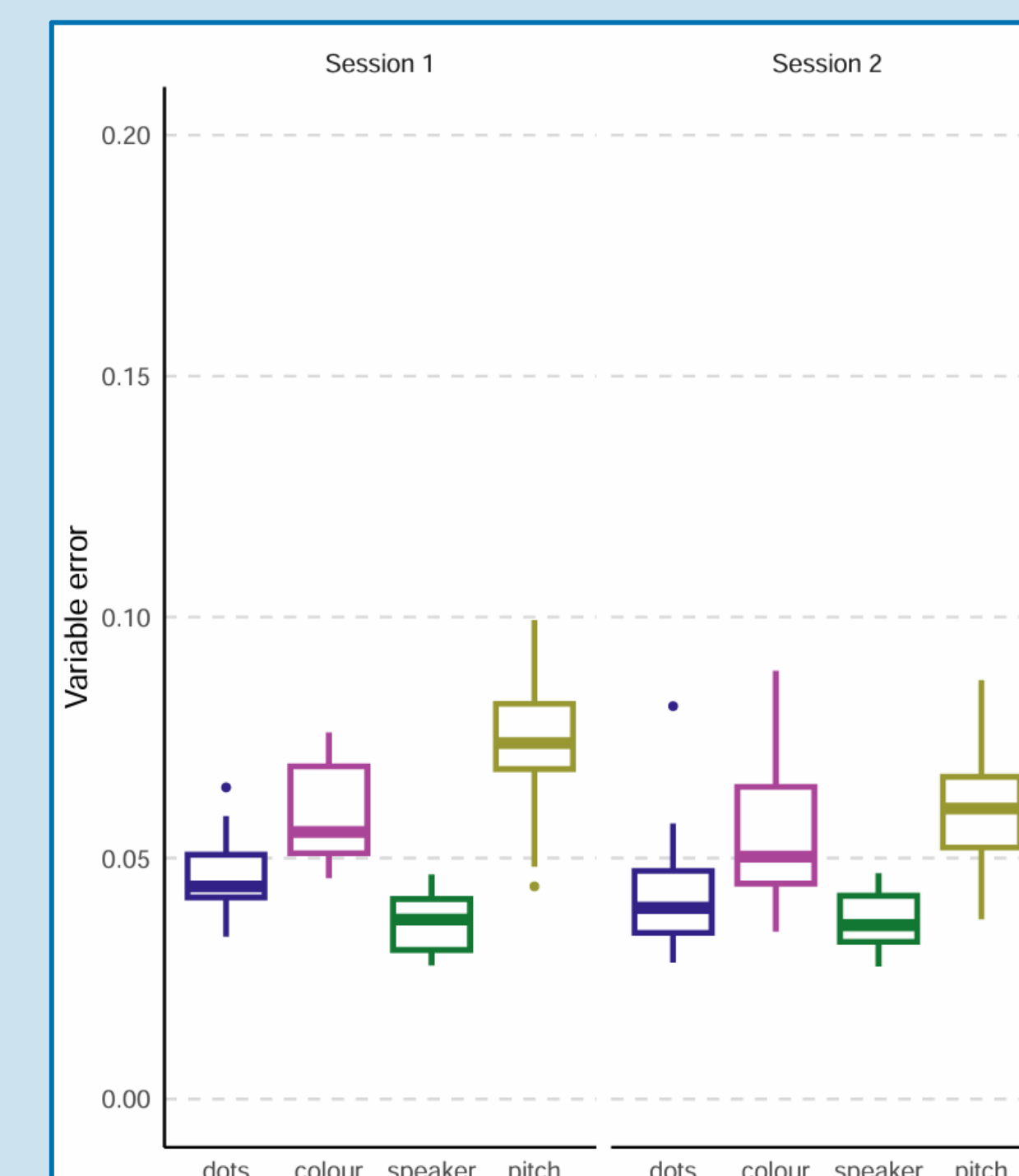
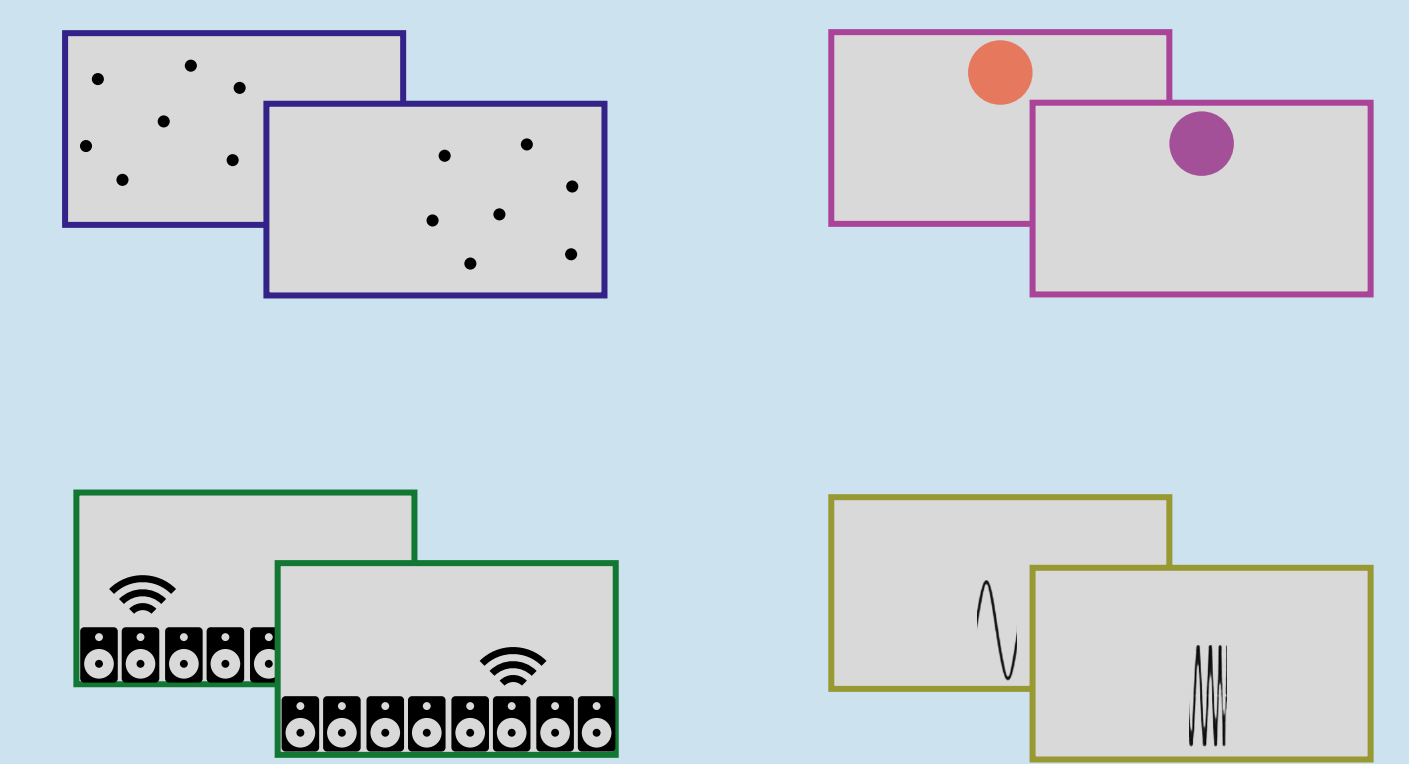
Ideally, all individual cues should have similar levels of variable errors (plot on the right), so that each pair of worst cue / best cue have VE ratios close to 1 (dashed line in the plot below).



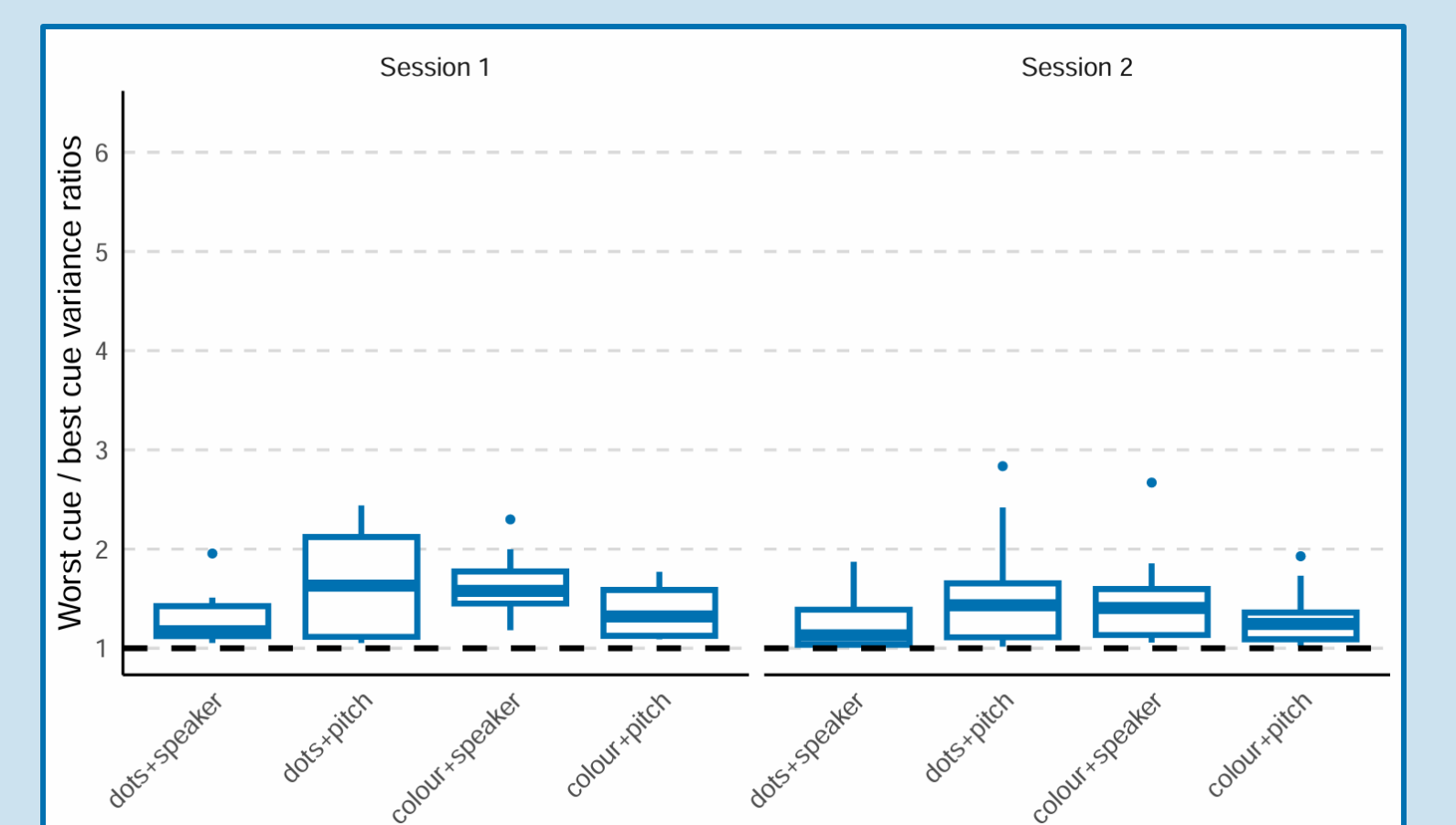
### Method 2: directional 2AFC task

In this task, 10 participants judged whether the second stimulus was higher or lower than the first stimulus in the dimension of interest: further left/right, more yellow/more pink, higher/lower pitched. Other than that, the staircase procedure is the same as method 1.

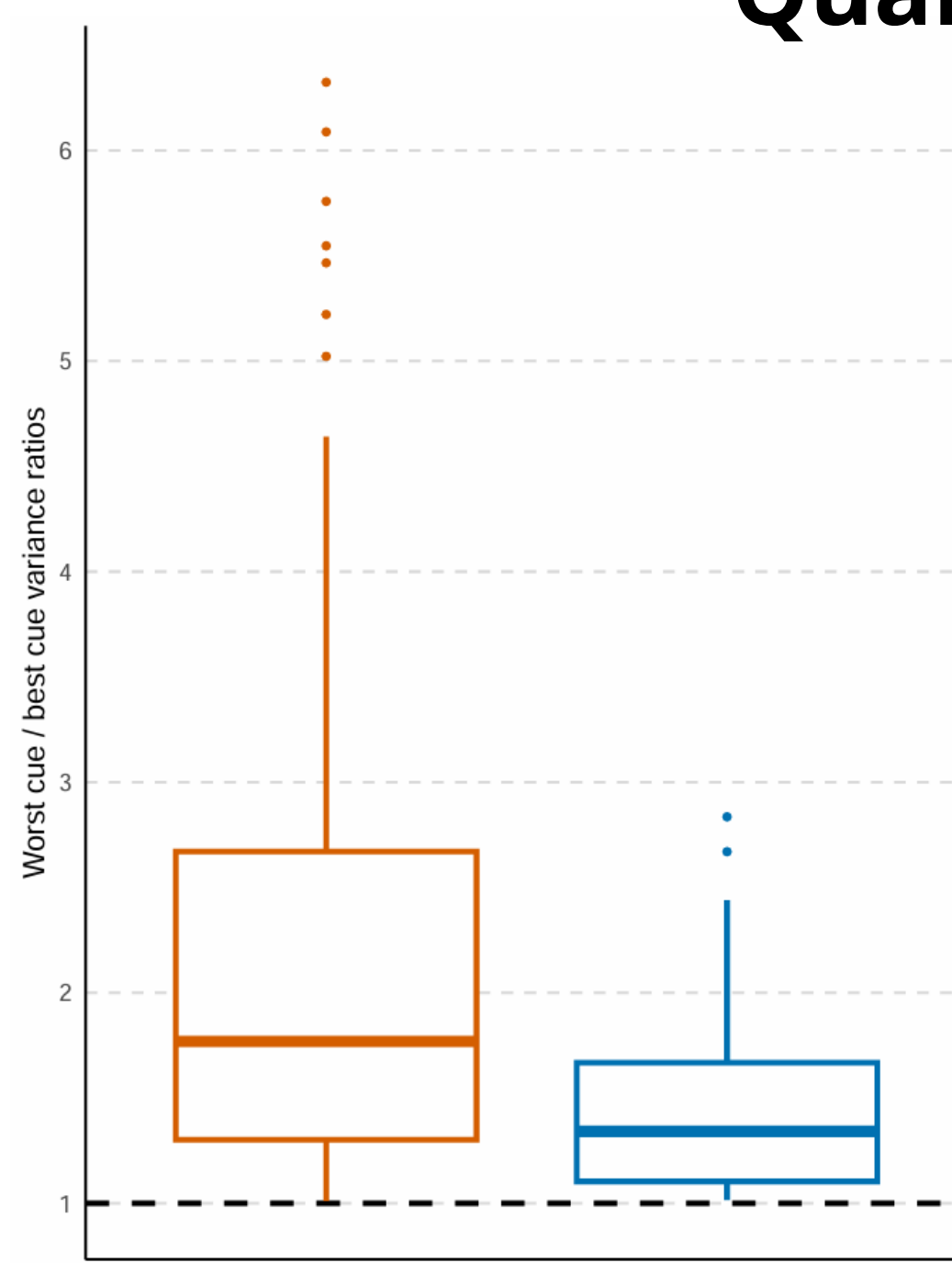
After some piloting, we also made the spread cue a bit harder (higher variance of the cloud of dots) and the pitch mapping easier (+25% thresholds).



We hypothesised that this *directional* task could be better suited to normalise cues given that the **cue-space mappings are also directional** (left - right). In these descriptive plots, this is reflected in the closer VE values for individual cues, and the worst/best ratios closer to 1.



## Quantitative comparison



As predicted, the worst cue / best cue VE ratios were lower in the directional 2AFC task ( $M = 1.45$ ,  $SD = 0.42$ ) than in the oddity task ( $M = 2.19$ ,  $SD = 1.21$ ), as confirmed in a linear mixed-effects model analysis accounting for the effect of the specific cue pairs and sessions as well as the random effect of participants,  $t(23) = -6.45$ ,  $p < .001$ .

The improvement was strongest for the *dots+pitch* pair of cues ( $M = -2.27$ ) and lowest for the *colour+speaker* pair of cues ( $M = -0.21$ ).

A ratio of 1 (dashed line) would correspond to perfect normalisation.

## Discussion

Task analysis is crucial to pick which normalisation task to match with the main multisensory task (here, directional fares better possibly because the multisensory task itself is directional)

Some of the spectacular improvement is also due to small changes we made to the dot and pitch cues (see Method 2 above).

If matching cues is a major concern, consider per-session or even 'continuous' normalisation (Rowland, 2025), as performance progressively improves with training. However, it could also change the cue-space mapping, making this approach unfit for some contexts like new cues.

## References

Ernst, M., & Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429-433.  
 Odegaard, B., & Shams, L. (2016). The brain's tendency to bind audiovisual signals is stable but not general. *Psychological Science*, 27(4), 583-591.  
 Rowland, B. (2025). Multisensory enhancement in detection and localization is reliable, robust, and consistent with an optimal computation. *IMRF 2025*, Durham.  
 Scheller, M., & Nardini, M. (2024). Correctly establishing evidence for cue combination via gains in sensory precision: Why the choice of comparator matters. *Behavior Research Methods*, 56(4), 2842-2858.  
 Wilbiks, J. M. P., Brown, V. A., & Strand, J. F. (2022). Speech and non-speech measures of audiovisual integration are not correlated. *Attention, Perception, & Psychophysics*, 84(6), 1809-1819.